

DOCUMENT RESUME

ED 174 651

TM 009 500

AUTHOR Anderhalter, O. F.
 TITLE Major Misconceptions About "Grade Equivalent" Scores.
 INSTITUTION Scholastic Testing Service, Inc., Bensenville, Ill.
 PUB DATE Feb 60
 NOTE 9p.; Official position paper of Scholastic Testing Service ; May not reproduce clearly due to colored background
 EDRS PRICE MF01/PC01 Plus Postage.
 DESCRIPTORS *Achievement Rating; *Achievement Tests; Elementary Secondary Education; Grade Equivalent Scales; *Grade Equivalent Scores; Norm Referenced Tests; *Norms; *Testing Problems; *Test Interpretation

ABSTRACT

The cautions needed for correctly interpreting grade equivalent scores are presented simply, for persons who are not statisticians. Four potential limitations of grade equivalent scores are summarized. (1) Comparisons across grades are of little value when the subject content is not the same; for example, sixth grade American history is not equivalent to seventh grade European history. (2) Because similar skills are not expected of students in different grades, grade equivalent results may be misleading. When a second grade student attains a grade equivalent of 5.0 on a second grade arithmetic test, it does not mean that he or she is capable of performing those arithmetic operations taught in fifth grade. It means that the student has performed as well as the average fifth grade student, on second grade work. (3) Being one year above grade level is more significant at the lower grades; significance is affected by the overlap in score distributions within and between grades. (4) Grade equivalent scores are not generally useful beyond the eighth grade. Percentile rank scores from norm referenced tests are recommended for interpreting test results. (GDC)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

MAJOR MISCONCEPTIONS ABOUT "GRADE EQUIVALENT" SCORES

by O. F. Anderhalter
Professor of Education
St. Louis University

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

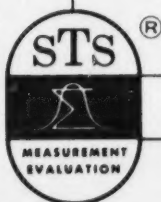
"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Scholastic Testing
Service

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

This article is an official position paper of
Scholastic Testing Service which was first
published in February, 1960.

SCHOLASTIC TESTING SERVICE, INC. • BENSENVILLE, ILLINOIS 60106



"Grade equivalent" scores are useful in tracking student growth on a single skill, such as reading comprehension. But these same scores can become very misleading when a teacher attempts to compare two or more grade equivalent scores *in different skill areas* for a given student. This position paper was prepared to help teachers see these hazards more clearly.

GRADE EQUIVALENT NORMS FOR ACHIEVEMENT TESTS

Most test users readily agree that raw scores obtained from achievement tests can be completely misleading when considered in themselves. Through training and experience, such users have come to accept the need for some type of norm when attempting to interpret the results from these tests. More often than not, test users have come to accept only one particular kind of norm for achievement tests, the "grade equivalent norm." When asked for a preference among the various types of norms possible, the usual response is in favor of grade equivalent norms, because "they are easier to interpret."

We at Scholastic Testing Service have a definite leaning *away* from grade equivalent norms, and our reason requires the addition of only a three letter prefix in the previous quote. Our feeling is that grade equivalents are "easier to *misinterpret*." Because of our convictions, we have prepared this position paper pointing out the limitations of grade equivalent norms.

To fully understand the limitations that will be discussed, it is necessary to have a general notion concerning the methods used in the derivation of such norms. Such a general notion can be gained with the background possessed by the typical classroom teacher. The following discussion describes the method ordinarily used to obtain grade equivalent norms. This discussion can serve as a background for the other discussions to follow.

HOW ARE GRADE EQUIVALENT NORMS OBTAINED?

As a first step in deriving grade equivalent norms for a test, the test is ordinarily administered to a sizable number of students in different grades. Perhaps a

test designed for Grade 4 would be given to a large number of students in each of Grades 1 through 8. Let us assume for illustration purposes that this was the case, and let us assume further that the test was given at midyear so that the actual grade placement of the students taking the test was 1.5, 2.5, 3.5, and so on up through 8.5 for eighth-grade students.

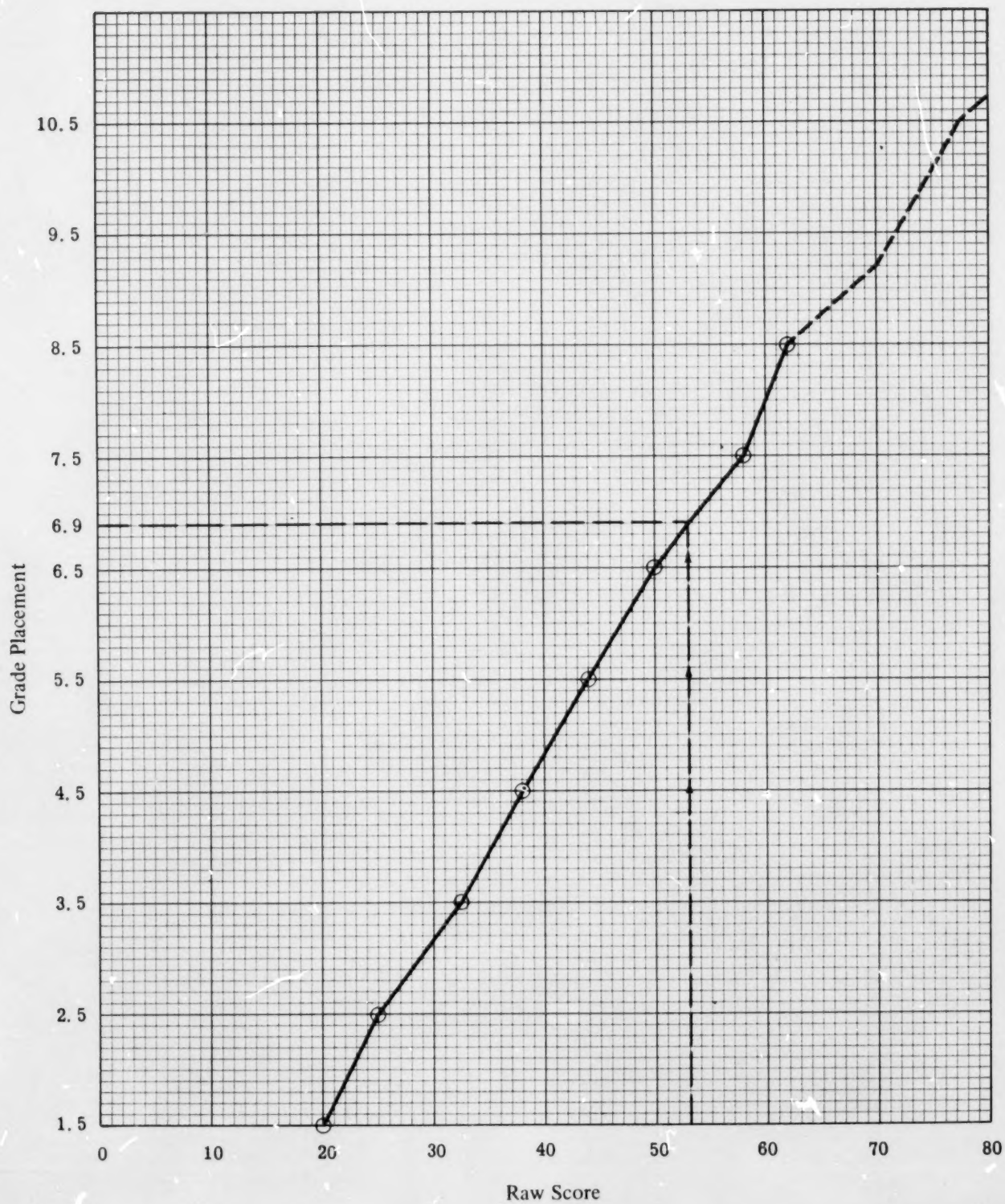
The next step involves finding some kind of an "average" score attained on the test by students in the various grades. More often than not, the "median" score is found for this purpose. The median is nothing other than the "middle" score, when all scores are listed from lowest to highest. Suppose, again, for illustration purposes, that the median score attained by first-grade students was 20; that attained by second-grade students was 25; and that for the other grades was as listed below:

Median Score On Test	Grade Placement
32	3.5
38	4.5
44	5.5
50	6.5
58	7.5
62	8.5

Such a table as this actually serves as the basis for converting raw scores to grade equivalents. Should a student in any grade attain a score of, say, 58, one can readily see that this is the same as that attained by the "typical" student at the middle of Grade 7. As a result, the score would be assigned a grade equivalent of 7.5. Of course, if the score were not specifically mentioned in the table — such as a score of 53 — it would be necessary to interpolate to find the grade

Figure 1

Typical Method Used in Deriving Grade Equivalent Conversion Scales



equivalent. By glancing at the table, it can be seen that a score of 53 would be opposite a grade placement of about 6.8 or 6.9. To make such interpolations easier, results are usually plotted on a graph. The graph in **Figure 1** illustrates such a procedure.

The dots inside the circles show how the median scores in the preceding table are plotted opposite the grade placement. Thus the 25 median score attained by Grade 2.5 is plotted opposite the grade placement of 2.5, and so on. After these dots are connected by straight lines, interpolation can readily take place. We show how the grade placement for a score of 53 can be obtained from the graph by the dotted line with arrows. Here the line ends nearest a grade placement of 6.9.

Aside from some "smoothing" of a graph such as this, you would now have the final conversion scale for changing all raw scores to grade equivalents. Publishers will, of course, usually read the values from such graphs and present them to you in table form.

WHAT ARE THE LIMITATIONS OF SUCH GRADE EQUIVALENTS?

With an understanding of the method used in determining grade equivalent norms, it is now possible to consider some of the resulting limitations. While not attempting to exhaust the subject, a few of the major limitations are discussed in the following sections.

For Many "Content" Areas, Grade Equivalents Make Little Sense

The comparisons "across" grades afforded by grade equivalent norms are of little value when content differs considerably from grade to grade. Suppose, as an example, that a school system teaches American History in Grade 6 and European History in Grade 7. If a test designed for Grade 6 were given to students in other grades, one can readily see what would happen. Students below Grade 6 would do poorly because they have not as yet studied the content covered in the test. Students in Grade 7 would likely do poorer than those in Grade 6 since they would likely have forgotten some of the material learned the preceding year. It would not be farfetched to expect the sixth grade to attain an average of, say, 50, while the seventh grade attained an average of, say, 40. This would lead

to such a contradiction as a score of 50 giving a grade equivalent of 6.5, while a lower score of 40 would give a higher grade equivalent of 7.5.

When a test is designed to cover the specific material in a grade, and when the content in other grades differs considerably, comparisons across grades as provided by grade equivalents make little sense. This is true of most of the "content" areas, but is true to a lesser extent in the "skill" areas.

Where Usable, Grade Equivalents Often Do Not Tell What They Seem To Tell

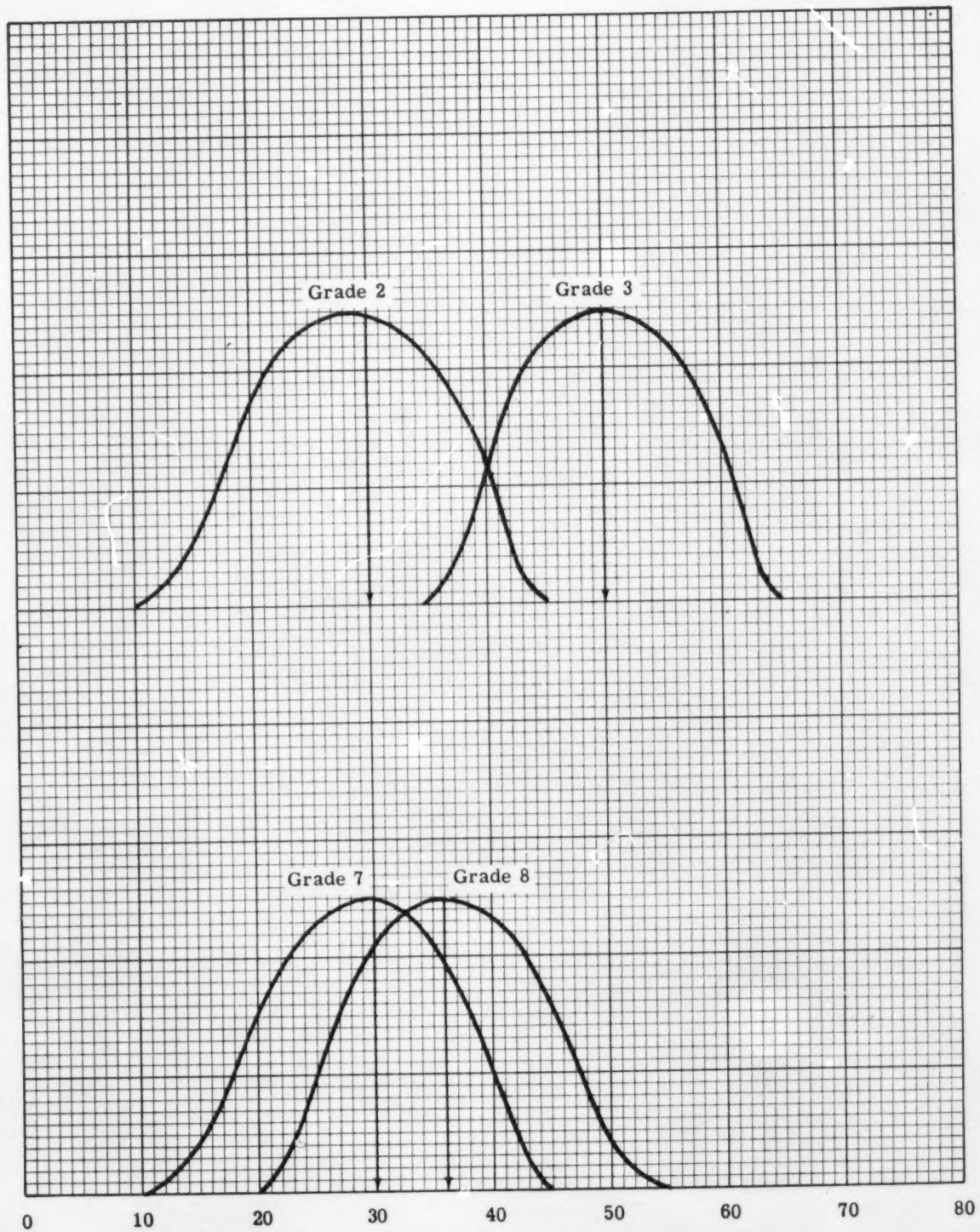
Where the content from grade to grade is sufficiently common to use grade equivalent norms — as in some skill areas at some grade levels — the norms often do not tell what they seem to tell. As an example, consider a second-grade student taking an arithmetic test designed for the second grade. Some second-grade student is likely to attain a grade equivalent of, say, 5.0, on such a test. What is the first impression one gains? That the student is capable of doing fifth-grade arithmetic! Stop one moment, however, to consider how one is able to say that a student can *multiply fractions* or perform other fifth-grade level work in arithmetic, *when the test did not have such types of problems in it.*

What has really happened is that, when obtaining the norms, the test covering second-grade material was given to students through the fifth-grade. Now, the second-grade student was able to perform second-grade work as well as the average fifth-grade student could perform *second-grade work*. This means the second-grader can add whole numbers with carrying, and similar second-grade level operations, as well as the typical fifth-grade student can perform these same second-grade level operations. This is quite distinct from being able to perform fifth-grade operations. Should the same child take a test covering fifth-grade material, one can readily see that the performance would be poor, and on the same type of grade equivalent scale, the student would likely attain a grade equivalent of 2.0 on the fifth-grade level test.

The important consideration is that no grade equivalent can take on direct meaning unless the test covers content of that grade level. Direct meaning is possible only over the range of content of the test. If a battery of tests covers material suitable to Grades 3 and 4, grade equivalents in this range have meaning; those outside this range do not have direct meaning.

Figure 2

Typical Overlap in Distribution of Scores



Grade Equivalents Often Mean Different Things At Different Points of the Scale

Is a year in advance of grade level *very good*? Is it just a little above average? Is a year below grade level *very serious*? Does it represent only a minor deficiency? Such questions as these cannot be given a blanket answer because the interpretation of such results often differs from test to test, and from grade to grade on the same test.

Consider first a nontechnical example. Suppose on a test covering only computations involving whole numbers, a second-grade student is a year above grade level. Suppose, also, on a similar test, a seventh-grade student is a year above grade level. Which of the two represents the more outstanding achievement? Obviously the second-grade student performance. This is true since the student is able to multiply, add with carrying, and perform other operations that have not as yet been taught! On the other hand, the eighth-grade group has been taught no new operations beyond those taught the seventh grade in relation to the content of the test. The content has reached the saturation point before the seventh-grade level. It is not unusual to find the average eighth-grade performance only a point or two higher than the average seventh-grade performance on such a test. In short, a year in advance at the one level means a considerable increase; whereas at the second level, it means only a slight amount above the average for the student's own class.

The same conclusion can be reached more directly by considering the amount of "overlap" in the distribution of scores attained in successive grades. Consider the graphical representation of scores made by students in Grades 2 and 3 and Grades 7 and 8 in **Figure 2**.

The graph for Grade 2 simply states that students in this grade made scores ranging from a low of 10 to a high of 45, with an average of 30 (the average represented by the line in the center with an arrow). In a similar manner, the graph for Grade 3 states that students in this grade made scores ranging from a low of 35 to a high of 65, with an average of 50. *Notice* that the highest score attained by any Grade 2 student (here, a score of 45) is still below the average of the Grade 3 students. If, therefore, a second-grade student did come up to the average of the third grade, this would represent outstanding achievement, since none in the original group tested scored this high. Here, a year in advance of grade level would be outstanding.

Now, consider the comparable graphs for Grades 7 and 8. It can be seen that the averages for the two grades are just five points apart. A considerable number of seventh-grade students attained scores higher than the average eighth-grade score — i.e., many seventh-grade students scored higher than 35, some scoring all the way up to 45. Here a year in advance is not as outstanding, since many other seventh-grade students have attained the same level.

This discussion points up the fact that the only way of determining whether or not a certain amount in advance of grade level is outstanding or simply just above average is through knowing the amount of overlap in the distributions involved. This, of course, holds for varying amounts, either above or below grade level.

The overlap in distributions such as is shown for Grades 7 and 8 is far more usual in tests than is that shown for Grades 2 and 3. On any test, some idea of the overlap can be obtained by noting how much difference in raw scores there is between two consecutive grade equivalents. Ordinarily, scores in a grade will range over forty or fifty or more points, *while the average difference between "averages" for two consecutive grades will be less than ten points!* The spread *within* a grade is usually far greater than the spread *between* grades.

Grade Equivalents Above 8.9 Rarely Allow Direct Interpretation

The previous discussions did not mention grade equivalents above the eighth grade on elementary achievement tests. The reason is that the usual method of deriving grade equivalents does not apply. If an elementary school test were given to students in the high school, it would not be unusual to find little or no improvement over eighth-grade results. More often than not, the ninth-grade class might do slightly better, with the tenth grade or eleventh grade doing about the same as the eighth-grade classes, and perhaps the twelfth-grade class doing even *poorer* than the eighth-grade class. This might be expected simply because high school students are no longer being formally taught the same content. Whatever the reason, it does occur.

As a result, higher grade equivalents are almost always obtained by "extrapolating" from the graph. The dotted line at the top of Figure 1 is such an extrapolation. One simply extends the line. *No two authors* will draw the line in the same manner. As a

matter of fact, it would be unusual for the same author to draw the line twice in the same manner. Some will follow such an arbitrarily set rule as setting a grade equivalent of, say, 11.0, equal to the top score attained by eighth-grade students.

Of importance is the point that grade equivalents above 8.9 are not obtained in the same manner as for lower grades. If a student attains a grade equivalent of say, 11.0, it ordinarily means only that the student is above the average for eighth-grade students, *not* that the student can do eleventh-grade work.

Intermediate Grade Equivalents Are Obtained By Interpolation

As illustrated in Figure 1, showing how a grade equivalent could be found for a raw score such as "53," interpolation is used for intermediate grade equivalents. This assumes "uniform growth" in the area from year to year. Using that illustration again, one can see that the average for mid-year sixth-grade students was 50, and that for mid-year seventh-grade students was 58. Consequently, one assumes that the average for 7.0 grade students would be midway between these, or at 54. Teachers know that growth in a subject is rarely "smooth" or continuous. A student may not understand a process for some months, and all of a sudden understands it completely. Thus, there might be no growth for several months, with a big jump at some point. The ordinary method of interpolating in grade equivalent scales does not allow for this possibility. For this reason, less emphasis should be given to small differences in grade equivalents. To place considerable emphasis upon a difference in grade equivalents such as 6.7 and 6.9 might be emphasizing only the assumptions involved in interpolating.

Summary

The preceding discussions point out serious limitations associated with grade equivalent norms. It can be seen that to properly interpret grade equivalents, one would need to know:

1. The exact range of content covered in the test.
2. The exact amount of overlap in distributions of scores from grade to grade, and,
3. The exact method of interpolation and extrapolation used by the authors in establishing the scales.

These limitations are well recognized by authorities in the field, and similar discussions appear in many standard textbooks in the field.

WHAT ABOUT PERCENTILE RANKS?

In percentile ranks, a student is compared only with other students at the same grade level. Thus, on a fourth-grade arithmetic test, a fourth-grade student is compared only with other fourth-grade students taking the same fourth-grade arithmetic test. The comparison is reported in a direct "ranking" type of scale. Thus, if a score is given a percentile ranking of, say, 80, it means simply that this performance was better than that of 80% of all students taking the test. This is almost directly comparable to the ordinary ranking system used by teachers on their own tests.

As such, percentile ranks are not subject to many of the limitations cited for grade equivalent norms. There is not the problem evidenced by the illustration concerning American History and European History tests, since the student is compared only with others in the same grade taking the same test. There is no problem of overlap of grades, nor of considering the range of content covered, since comparisons relate only to a single grade. The same holds for the other limitations of grade equivalents discussed.

The "Technical Recommendations for Achievement Tests" issued jointly by the National Council on Measurements Used in Education and the American Educational Research Association stresses the advantages of percentile norms in various discussions. The publication, for example, specifically states that percentile norms should be provided to help interpret scores whenever grade equivalents are provided. It further notes that percentile norms should be used unless there is some strong reason for using other types of norms.

Scholastic Testing Service, in accordance with the above discussions, places strong emphasis upon the use of percentile norms for its achievement tests. It does "bow" to public demand to the extent of providing grade equivalents for specific scores on certain tests. The users are cautioned of the possible misinterpretations, however, and are urged to use the percentile norms in conjunction with the GE's.